

## University of Colorado, Boulder CU Scholar

---

Undergraduate Honors Theses

Honors Program

---

Spring 2011

# Harm, Feelings, and Wrongness

Michael Lind

*University of Colorado Boulder*

Follow this and additional works at: [http://scholar.colorado.edu/honr\\_theses](http://scholar.colorado.edu/honr_theses)

---

### Recommended Citation

Lind, Michael, "Harm, Feelings, and Wrongness" (2011). *Undergraduate Honors Theses*. Paper 620.

This Thesis is brought to you for free and open access by Honors Program at CU Scholar. It has been accepted for inclusion in Undergraduate Honors Theses by an authorized administrator of CU Scholar. For more information, please contact [cuscholaradmin@colorado.edu](mailto:cuscholaradmin@colorado.edu).

# **Harm, Feelings, and Wrongness**

Michael Lind

Department of Philosophy

Thesis Advisor: Brian Talbot

Dr. Brian Talbot  
Philosophy Department

Dr. Robert Hanna  
Philosophy Department

Dr. Leaf Van Boven  
Psychology Department

Archival Copy

4/11/2011

## **Abstract**

While feelings and emotions are quite important in daily life, their role in morality has often been underplayed. Although there have been some ethical theories, such as David Hume's sentimentalism, that place feelings as an important cornerstone of morality, feelings are often regarded as motivators for moral actions at best and destroyers of moral character at worst.

I will argue that feelings play an important role in morality via the nature of harm. Feelings, when properly defined, are present in all cases of harm as either actual or possible negative experiences. A feeling is a non-localizable mental phenomenon that has an affective experience. This experience can be crudely classified as positive or negative; feelings such as pain and anxiety belong in the negative category. It is these negative feelings which cause harm to be bad for the one who experiences it.

Harm is often defined in either a comparative or a non-comparative account. The comparative account of harm is fatally flawed since it cannot account for overdetermined harms and difficult comparison cases, among other things. As such, the non-comparative account is to be preferred. Under the non-comparative account of harm, death is not harmful, and neither is anything else that cannot be experienced. Thus I will argue that only things that can be both experienced and felt in a negative fashion can be harmful for a person.

At least some wrong acts are wrong because they are harmful. Everything that is harmful is dependent on feelings, so some wrongful acts are dependent on feelings. While harm is not the only factor that can cause an act to be wrong, since there are acts that are harmful without being wrong, harm plays an important role in morality.

## Feelings

I have found that most of the current definitions of feelings and emotions are inadequate. Some definitions focus solely on their physical and behavioral aspects, ignoring the affective, subjective component entirely. Other definitions equate feelings with emotions, something I will argue about later in the paper. If I am to argue about how feelings relate to harm, then a new, or at least revised, definition of these concepts is needed.

I will start by defining what a feeling is. I will first lay out an explanation, then compare feelings to other mental phenomena, and then finally give a concise definition. One necessary condition for a feeling is that it must be a mental phenomenon that is causally dependent on a physical stimulus. While feelings are dependent on physical stimuli, there are usually different stimuli for different feelings. An example might be that the feeling of sadness affects certain receptors in the brain, whereas the feeling of pleasure affects other neural pathways. The exact stimuli in question are more the interests of neuroscientists than philosophers. For this paper, it is just important to note that feelings have corresponding physical pathways, but different feelings have different pathways.

Despite their causal reliance on physical stimuli, feelings are not reducible to their causes. This is because feelings have a subjective “felt” experience that cannot be completely explained by physical descriptions.<sup>1</sup> For example, a feeling of pain registers the mental “ouch” phenomena while a purely physical sense of pain would register the firing of c-fibers, but not the “ouch” feeling one normally associates with pain. Since this is a thesis in philosophy and not psychology, I will focus on the experiential aspects of feelings and not their corresponding physical processes. Unfortunately the experiential nature of feelings makes them notoriously difficult, if not

---

<sup>1</sup> One might consider feelings a type of qualia, if one wishes. It should however be noted that I will neither attempt to argue against or for the existence of qualia in this paper.

impossible, to perfectly define in words. Because of this, I will have to assume that the reader knows what I mean when I refer to subjective qualities of feelings such as sadness or fear.

What differentiates feelings from the other phenomena is the particular evaluative nature of feelings. When I say feelings are evaluative, I mean that feelings are evaluative judgments, since they classify things in a certain way. This requires more elaboration. First of all, since not all feelings have moral worth, I will use the terms “positive” and “negative” to refer to the evaluative judgments made by feelings. In this usage, the feelings are being evaluated as positive or negative based on how the feeling affects the one experiencing them and not on the utility of the feelings. Fear can be quite useful in motivating people to avoid dangerous or harmful objects, and thus it can have a high utility. However, the subjective experience of fear is always negative in virtue of it being an unpleasant feeling. Fear may have instrumental value, but it doesn’t have intrinsic value. This can be evidenced by the intuition that unjustified paralyzing fear is always bad for the person who experiences it.

The terms “positive” and “negative” are also a bit of a simplification. Even feelings that are all negative (pain, sadness, and anxiety for instance) have different subjective experiences. This may be analogous to the difficulty found when one tries to verbally explain the differences between color sensations. Red is different from yellow by virtue of it looking different, just as anxiety is different from sadness by virtue of it feeling different. One might argue that anxiety and sadness can also be distinguished by physiological measures such as heart rate and stress, but this only captures the physical aspects and not the experiential aspects of the feeling. It would be like measuring color sensations scientifically by wavelength; a spectrometer can tell us how many nanometers apart the colors might be, but it cannot describe the sensation of red or yellow in a meaningful way.

Feelings can be distinguished from memories and thoughts by analyzing the types of evaluations they make. Memories can have affective evaluations just like feelings do, but this is because the evaluative aspect of memories is parasitic on the evaluative nature of feelings. Memories that are “good” bring up positive feelings, and memories that are “bad” bring up negative feelings. In terms of pure recollection, memories cannot be good or bad if one has no feelings about them. Thoughts may be evaluative judgments as well, but these are more nuanced than the affective judgments made by feelings. Feeling-based judgments are distinct from “thoughtful” judgments in that judgments of feelings are based on the subjective experience of the feeling as opposed to cognitive appraisals that are made by thoughts.

Although the evaluative aspect of feelings differentiates them from some mental phenomena, it does not separate them from sensations, which are also evaluative. Robert Solomon suggests defining a feeling as a “non-localizable sensation”.<sup>i</sup> Although this is not entirely accurate, since feelings do not always have sensations, it is a workable definition because in the real world feelings do have corresponding sensations. One example that might illustrate the possible differences between feelings and sensations is the sensation of itchiness. The sensation of itchiness has both a physical dependent cause and an associated subjective sensation that is distinct from its cause. And while it can be referred to as a feeling in casual language (“feeling itchy”) it is not a true feeling since itchiness has a corresponding physical location. The same cannot be said for the feeling of sadness; sadness has no corresponding part in the body whereas itchiness seems to have a particular location, usually on the surface of the skin.

The nature of pain suggests that although the feeling of pain often has a corresponding sensation, it need not. Pain may or may not have a physical location, depending on whether one

is talking about physical pain or psychological pain. This distinction might be somewhat arbitrary, as it has been found that neurological scans of psychological pain are quite similar to the scans showing physical pain. (MacDonald & Leary 2005).<sup>ii</sup> Nevertheless, I will be using examples where physical pain can actually be “pinpointed” at a physical location. In most cases, the sensation of pain is concentrated around the area where the physical damage is. If one’s hand is hurt, then the sensation of pain is sensed to be in one’s hand. In contrast, the feeling of psychological pain does not match up with any part of the body, though the physical cause of the psychological pain is brain based. The pain of being socially ostracized still carries an experience of pain, but it does not have a physical counterpart it could match up with. It also requires a cognitive component as well; if one does not think (or simply doesn’t care) that one is being ostracized, then there will be no psychological feeling of pain associated with the idea.

There are other pain related phenomena that suggest differences between feelings and sensations of pain. For instance, there are times where the pain has a location, but it does not match up with the damaged area that is sending the signal. The pain one feels when suffering a heart attack has a particular location, but it is located in the left arm instead of the heart. Another example would be the phantom limb phenomenon, where the person experiences pain that seems to be located in a limb that is no longer attached to the person.<sup>iii</sup> It has also been shown that the feeling of pain and the sensation of pain can be separated by scientists in the lab. One experiment used participants who were hypnotized and then had their hands placed in painfully cold water (Hilgard 1975).<sup>iv</sup> The subjects reported that they sensed pain but were not bothered by it. It seems that in this case the physical sensation of pain had been divorced from the subjective experience of it. It should be noted however that while the experience and sensation of pain are not necessarily connected, they are intimately tied together in everyday life. Feelings often do

have corresponding sensations, but the sensations are not actually necessary, just very common. While sensations can be thought of a physical analogue to feelings, feelings are the part that does all the work when both feelings and sensations are involved. For instance, the sensation of pain only feels bad if one would rather be rid of it.

The definition of a feeling can be summarized as follows: Feelings must 1) be causally based on physical stimuli, 2) have a subjective quality that is not reducible to a physical explanation, 3) contain an evaluative judgment that is either “positive” or “negative”, and 4) lack a corresponding physical location.

### **Emotion**

Having hopefully defined what feelings are (and how they differ from sensations and other mental phenomena) I will now move on to the discussion of emotions. Emotions are just as difficult to describe as feelings, if not more so, so while my definition should capture most emotions, there may be a few rare cases that defy categorization.

Part of the difficulty in defining emotions and feelings is due to the common usage of terms. In casual speech, feelings and emotions generally have the same names. One can feel sad, but sadness can also be an emotion. The same goes for the feeling of anger versus the emotion of anger, as well as many other examples. However, there are a few rare cases where feelings and emotions do not share the same name. For instance, pain is a feeling but it is generally not thought to be an emotion. There are also feelings that have no correspondence to emotions whatsoever. Being nauseous is a feeling, but there is no emotion that corresponds to that feeling.

Before I move on to my own definition of emotion, I will analyze and critique some of the current theories of emotion. These will include the feeling theories and cognitivist theories.



The feeling theories of emotion suggest that emotions are reducible to feelings. Feelings, as William James defines them, are “changes in physiological conditions relating to the autonomic and motor functions.”<sup>v</sup> Feelings are based on behavior, as James states “we feel sorry because we cry, angry because we strike, afraid because we tremble, and [it is] not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be”.<sup>vi</sup> Feelings (and emotions, since James classifies them as the same thing) can be differentiated based on the different body responses that cause them.

Although James’ theory may have some merit since behavior can influence emotions, there are still some problems that need to be addressed. One issue with the James-Lange theory is the idea that emotions are distinguished by their unique physiological processes. This is not true for emotions, since some emotions have the same biological expression but are still recognized as distinct from each other. The best evidence for this objection is the Schachter-Singer study in 1962.<sup>vii</sup> In the study, some participants were injected with epinephrine while others were given a placebo. The participants were then placed in a room with a confederate, who either acted angry or giddy and euphoric. Since James’ theory asserts that it is physical changes that lead to emotions, then the participants in both groups should experience the same emotion since they have the same physiological reactions. However, that did not occur. The majority of the participants in the “angry” condition proceeded to become angry and act out typical angry behavior. The same was true for the “happy” condition; the behavior and feelings of the participants modeled those of the actor, not the physiological sensations the participants were having. This study and other ones that have obtained similar results<sup>viii</sup> suggest that James’ hypothesis that feelings are synonymous with emotions was incorrect, since the physiological

states of the participants were very similar but their emotions and behaviors were widely different.

Cognitivist theories state that there are cognitive components that distinguish emotions from feelings, though what exactly those components are depends on the particular theory. Many philosophers, including Robert Solomon, believe that emotions are judgments which can express propositional attitudes. While this may be true it is only a necessary and not a sufficient condition for defining emotion. One objection, supported by Deigh (1994)<sup>ix</sup> argues that if emotions are just propositional attitudes then entities incapable of language, such as animals, cannot have emotions. And yet it does seem that some animals (particularly the great apes) do express emotions; indeed some chimpanzee expressions closely resemble those of humans. So either animals can't have emotions, or there is more to an emotion than just propositional content.

My personal definition of an emotion is a modified cognitive view. Emotions do have propositional content, though what is important is their intentionality towards things. Emotions have feelings as well, although they are not just feelings. This is because some emotions can have the same physiological characteristics (and in some cases even the same qualitative feelings) but remain distinct in their intentions and expression. Emotions are not just judgments either; even though emotions do necessarily involve intentions towards objects this definition alone does not distinguish them from thoughts and mental phenomena that can have intentions. As the cognitivists suggest, emotions are evaluative, but I will argue that this is based on the feelings an emotion has and not the properties of the emotions themselves. My definition of an emotion is something that has two major components, a feeling and an intention towards an object.

The notion that emotions involve feelings seems to be a contingent natural fact rather than a necessary law. Thus some of my arguments in favor of emotions having feelings will be naturalistic. First of all, emotions can have physiological and behavioral expressions that are exactly the same as those had by feelings. This by itself is not enough to prove anything, since they are only contingently related; this would be true even if the relationship was a natural fact about human and animal psychobiology. If one takes a behavioral approach to feelings and emotion, it is impossible to distinguish an emotion from a feeling, which help explains why feeling theorists regarded them as the same and behaviorists ignored them as a subjective phenomenon altogether. It will be easier to show the importance of feelings to emotions by imagining what emotions would be like without feelings.

Under the current definitions of feeling and emotions, “feeling-less emotions” would be contradictory. Although they are not emotions per se, it might be easier to call these entities empty emotions. The contradiction arises because feelings have two effects in an emotion; they create the subjective experience and they motivate certain behaviors while inhibiting others. To illustrate this contrast between emotions and empty emotions, let’s say I am experiencing the emotion of anger towards someone who was stepped on my toe (whether the emotion is justified in this example is irrelevant). A normal emotion of anger has all the basic parts: I am mad at the person *because* he stepped on my toe and I have some motivation towards certain kinds of behavior. A person with an empty version of anger would not experience any feeling of anger whatsoever; this also means there would be no physiological reaction commonly associated with the feeling, such as the activation of the sympathetic nervous system. More so, the “angry” person would not have tendencies towards any aggressive actions at all, assuming there aren’t any other motivational factors at work. Without feeling, emotions don’t have any motivational

force at all, and since emotions are well known for their motivation, feeling and emotions must be intimately connected.

Emotions also have intentions. What distinguishes emotions from one another are their formal objects, not their feelings. This can best be illustrated by looking at two emotions, the emotion of indignation and the emotion of anger. An example of anger could focus on the fact that my bike was stolen. The anger would be directed at the person who stole my bike.

While anger is focused on being personally offended, indignation's offense is based on the principle of the matter, and not personal reasons. An example of indignation could concern someone's theft of a bike that I was about to put up for the garage sale the coming day. The bike had no monetary or sentimental value, so it's not the loss of the bike I am indignant about. Rather it is the principle of the matter: I could care less about the bike itself; it is the fact that my bike had been stolen that infuriates me. When one is indignant, one is offended by the breaking of principles and not the act itself. However, anger and indignation are not mutually exclusive; one could be angry and indignant at the same object (and thus mad at both the offense and the offender).

This particular example serves to reinforce an important point: emotions are distinguished from each other by what they are directed at instead of their qualitative feeling. As it happens, the emotion of anger and the emotion of indignation could both have the feeling of anger (both emotions might feel exactly the same). The fact that indignation and anger are so qualitatively similar explains why they are often confused as being different variations of the same emotion, with indignation being considered a special version of anger.

One possible problem with this definition of emotion is that emotions have intentions that can seemingly be directed at non-existent objects. This is not a particular problem with my

definition; it affects all definitions of emotion that require intentionality (which are mainly the cognitive theories.) One example would be an emotion of fear that seems to be directed at the monster under the bed, despite the fact that no such monster exists. Although problematic, this issue has multiple solutions. One solution would simply be to bite the bullet and suggest that one can indeed have intentions towards imaginary objects. While this is a possible solution to the problem, it would require me to go over many issues that while interesting, are not all that relevant to the purpose at hand. This issue can be circumvented for now by only using examples of emotions that are directed at real objects.

One might object that there are objectless emotions, which defies my definition of emotion since it lacks an intentional object. Objectors might try and use an example of objectless fear; the victim experiences the “emotion” of fear, but there is no object to be afraid of. It is not that the person cannot name the object of the fear or that the fear is unknown; there is simply no object at all. Scientific studies have shown that it is possible for one to be afraid of nothing in the sense that one is afraid but there is no exact object of fear. For example, some studies investigated a stimulated part of the hippocampus in people and found that the subjects experienced an intense feeling of fear that had no object (there was nothing they were scared about).<sup>x</sup> While the direct brain stimulation caused the fear, it would not be correct to say the object of the fear was the brain stimulation.

This objection can be resolved by pointing out that objectless emotions are really just feelings, since the feeling component of the emotion is still intact, just without a specific intention. Going back to the objectless fear, it would be an instance of a feeling of fear, but not an example of the emotion fear. It is easy to understand how this confusion came about if one remembers what I said about the nomenclature for feelings and emotions earlier in the paper. The

emotion of fear can be confused with the feeling of fear in causal usage (since they have the same name), but the two things are actually different. While it is doubtful causal usage of the words can be changed to reflect this distinction, philosophers would do well to make sure the differences between the two are reflected in the terminology.

### **Harm**

Having discussed the differences between emotions and feelings, I will now go on to focus on two points: the relationship between harm and feelings and the relation between harm and wrongness. While not everything that is wrong is based on harm, everything that is harmful is based on feelings. Harm is morally significant, but it is not the only morally relevant thing. My main argument can thus be outlined as follows:

- i) The notion of harm is based on feelings
- ii) The Western notion of wrongness is partly dependent of the notion of harm
- C) Therefore the Western notion of wrongness is dependent on feelings, at least in part.

I restricted my argument to Western morality because some moral notions of wrongness do not depend upon harm. Haidt (2007) suggests there are five foundations of morality: harm, fairness, ingroup loyalty, authority and purity.<sup>xi</sup> The last three are less important in the Western world, so I will exclude them. While I do believe that all five foundations can arguably be based on feelings, I will only focus on harm for this paper.

I will attempt to show that wrongness is dependent on feelings, but I am not advocating consequentialism (in particular utilitarianism) although some consequentialists would accept my conclusion since the definition of wrong for consequentialists is having the maximum amount of pain (which as I have already discussed is a feeling). Before I discuss wrongness, I will define a notion of harm.

I will start by elaborating on the definition of harm. Hanser (2008) suggests there are two main accounts of harm.<sup>xii</sup> One is the comparative account, and the other is the non-comparative account. For someone to be harmed on the comparative account is for that person to be placed into a worse state than he or she would have been otherwise. For example, someone who is hit on the head is harmed because that person is worse off than she or he would have been otherwise. That person now has a headache, whereas that person would not have had one if he or she had not been hit on the head. The same example could be used for the non-comparative state based account of harm. The person is placed into a bad state, since he or she now in pain. Some states may arguably be bad, but pain being at least a *prima facie* bad is relatively uncontroversial.

The comparative account of harm appears to be quite plausible at the outset. In simple commonsense cases, it works quite well. The headache case mentioned above is one example. However, there are several features of the comparative account that are problematic. Some of these problems concern absurd cases of harm, harm that cannot be compared, and figuring out what sort of cases should be compared.

At first it may seem that if events harm someone by making that person worse off, one can be harmed simply because that person did not receive a benefit. Thus it might appear that all failures to benefit are counted as harms on the comparative account. However, Hanser argues that that is not really the case. Many failures to perform a benefit are really non-events, and thus not actual events. One could retort that non-events are actually particular species of events, but I will not look into this possibility here. If one does not classify non-events as a particular type of event, then not all failures to benefit are harms. A person would have been better off if that person received the benefit, but because the would-be benefit never happened, it is a non-event and thus cannot be viable for the comparable account. All events still count in the comparative

account though, so if one actively prevented another from receiving a benefit, that would be a harm.

Even if non-events do not count as events, the reliance on events can still lead to some absurd cases of harm. For example, suppose Suzie has been given a bit of money to buy a toy. As she is walking down the street, she notices a plastic bag floating in the wind and chases after it, being a resolved environmentalist. In doing so, she fails to notice the first toy store she passes by and enters the first one she actually notices. In the store, she buys a toy that she likes. However, it just so happens that had she noticed the first store, she would have bought a toy that she enjoyed even more and thus would have been happier. So according to the comparative account of harm, Suzie is harmed by noticing the plastic bag since it lead her to be in a worse state than she would have been otherwise. Had she not noticed the bag she would have be happier, so on this account it would seem that the bag harmed her or at least led to her being harmed, since the event made her worse of then she would have been before.

This problem can be summed up by stating that when there is an event involved, someone is harmed if the actual event does not make them better compared to a relevant possibility. This leads to the conclusion that if benefits are not maximized due to a certain event, that person has been harmed by the occurring event. And yet it seems that although people can be harmed even if things end up going well for them, the conception of harm should not be so broad as to suggest that people have been harmed just because the actual scenario was not as good as a possible one.

A second problem with the comparative account is that it cannot explain harm in cases where no other alternatives are possible. McMahan lays out the example of the incurable patient.

“A person dies at age twenty from a congenital, genetically based condition. This condition is invariably fatal and no one who has had it has ever lived beyond the age of twenty. There is no cure and none on the horizon, no treatment to extend the victims’ lives, and no



recorded instance of regression, remission, or spontaneous recovery. Although the condition causes premature death, it does not cause significant pain or disability during the victim's life."<sup>xiii</sup>

McMahan goes on to suppose that the fatal genetic condition is essential to this person's identity.

If the patient was born with normal genes, he or she would have been a different person. Since it is controversial whether genes are essential to a person's identity, this may have to be assumed for the sake of argument. If these conditions are accepted, then the comparative account has another issue.

In the incurable patient example, the patient cannot be harmed by the disease on the comparative account since there was no possible world in which the patient could exist and not have the disease. Even though most people would be tempted to agree that the disease was harmful to the patient, the comparative account cannot explain the harm when there are no other possibilities.

While the two objections above are somewhat problematic, there is another problem that makes the comparative account an inadequate model for harm. This problem is based on criteria; exactly what sort of comparisons should be made between the actual state of affairs and the possible ones when determining how one could be worse off? McMahan calls this the metaphysical problem. Although this problem is pertinent in all comparative instances of harm, they are most perplexing in cases of death as a supposed harm, so I will focus on some cases concerning death.

I will assume in this paper that there is no afterlife, and that when one dies, one ceases to exist. And if one does not exist at death then it makes well-being comparisons between the living and the dead difficult, since the dead have no level of well-being to speak of. Rather death is usually thought to be harmful on the comparative account because it consists of a deprivation of goods; a person is deprived of the goods of life when he or she dies.

Even if someone is harmed, sometimes it is difficult to figure out what comparative scenarios should be included and which should be excluded. If the comparative account's purpose is to figure out when someone has been harmed by comparing different possibilities, which possibilities are supposed to be compared? In the case of death, it would be absurd to compare a person's untimely death with a scenario that involves that person being immortal. If that was the case, then every person who died would have been harmed by death just because they were not immortal. It is also quite intuitive to exclude possible but incredibly implausible situations as well. In the case of the incurable patient's death, it would not be acceptable to suggest that the patient was harmed since the patient would have fared better if he or she was cured by benevolent space aliens. So context must play a key role in determining which possibilities should be included in the account and which should be excluded.

Some comparative scenarios appear easier to explain than others. McMahan brings up the case of the pedestrian: "A person is absorbed in his own thoughts and steps absentmindedly off the curb into the path of a bus. He is killed instantly by the impact."<sup>xiv</sup> An obvious alternate possibility for the pedestrian would have involved him not getting hit by the bus and killed. But McMahan brings up another just as likely possibility; instead of getting hit by the bus straight on the pedestrian could have just suffered a glancing blow instead. This blow could cause a variety of injuries; it could be something as minor as a bruise to permanent brain damage, depending on the strength of the impact. If these possibilities are about equal in chance (or close to equal) for the pedestrian, then it becomes more difficult to discern whether the pedestrian was harmed or not. If he receives a bruise from the bus, then he is worse off than if he missed the bus altogether but not as bad off as he would have been if he got hit harder. Since any contact with the bus will make him worse off than he was previously, a proponent of the comparative account might

simply be able to say that any contact with the bus is harmful. However, this explanation runs into another difficulty.

Even if the pedestrian case can be more or less resolved, there are still other metaphysical issues at hand that the comparative account has trouble addressing. There is yet another problem, which McMahan calls the overdetermination problem. An example provided by Hanser provides a good example of overdetermination:

“Suppose that a criminal wants to steal from S’s store. Since the burglary will go more smoothly if S is not present, the criminal hires some thugs to break S’s legs the day before the proposed crime. When the thugs arrive at S’s home, however, they find that the local loan shark is already there breaking S’s legs. So they abandon their plan.”<sup>xv</sup>

In this case, one would want to say that the loan shark has harmed S greatly by breaking his legs. The non-comparative account can fit this intuition; in the non-comparative definition of harm, S is put into a bad state of excruciating pain by having his legs broken, so S is harmed. The thugs are irrelevant to the actual harm, even though they would have broken S’s legs had the loan shark not. This is not true for the comparative account of harm. Had the loan shark not broken S’s legs, the thugs would have broken his legs shortly after. Thus, under a strict reading of the comparative account, the loan shark had barely harmed S at all, since S’s legs would have been broken soon after by the loan sharks anyway. One can modify the comparative account to get around this issue, but as Hanser points out this only leads to other problems. For example, if one wants to compare the actual situation to the situation in which neither the loan shark or the thugs break S’s legs, then one could formulate the comparative account as follows: “someone suffers harm if and only if there occurs an event such that he would have been better off, for some period of time and in some respect, had neither that event nor any relevantly similar event occurred.”<sup>xvi</sup> This leads back to a few of the issues that I already covered. One problem is what counts as a relevant harm and what does not. This leads us back into the metaphysical problem

again. This also creates odd comparisons when death is involved since the subject's death now needs to be compared to cases where he doesn't ever die (not dying is relevantly different from dying). A good account of harm should not have to compare actual cases of death against immortality, and this revised comparative account fails spectacularly by requiring such comparisons.

Unlike the comparative account, the non-comparative account of harm relies on bad states to explain harm. On this account, one is harmed whenever one is put into a non-comparatively bad state. Although it is called the non-comparative account, this theory may actually use implicit comparisons in determining what things are non-comparatively bad. Hanser says that "when a proponent of the non-comparative account says that states of impaired functioning are 'non-comparatively' bad, he means rather that it is bad for a person to be in such a state regardless of whether a better state was ever a genuine alternative for him"<sup>xvii</sup>. For instance, if pain is a bad state and someone suffers pain, then the sufferer has been harmed. This is true whether there was another scenario in which the sufferer could have avoided the pain or not.

In my opinion, non-comparative bads must meet a few criteria to actually be considered non-comparative bads. First of all, non-comparative bads must be intrinsically bad or necessarily lead to an intrinsic bad; if a candidate's "badness" can be explained more fundamentally by another non-comparative bad, then it is not really a non-comparative bad. For instance, blindness prevents one from experiencing certain pleasures, particularly aesthetic ones. It also makes one more likely to experience pain, either physical pain like stumbling or perhaps psychological pain from not being able to do certain activities that most people can. Sickness can be explained in a similar way. Sickness is painful, or at the very least, preventative of some

pleasures. Pain is different because it doesn't have a further justification; pain is bad because it feels bad.

That being said, the painfulness of specific things is contingent rather than intrinsic. The human body is programmed to respond to certain stimuli negatively. Placing one's hand on a hot stove is painful, though theoretically it could have been otherwise. The feeling of pain itself however is intrinsically bad because it is not reducible. This can also be shown by the fact that no one ever wants to feel pain for its own sake. Even masochists like some types of pains only because they cause pleasure as well. Masochists also only pursue certain types of pain; it is highly unlikely that a masochist would enjoy having a tax audit or waiting for hours at the DMV. Pain is intrinsically bad as a natural fact; the badness of pain cannot be reduced to anything else whereas the badness of many other supposed "intrinsic" bads can be reduced to either possible or actual feelings of pain or other negative feelings.

An adequate non-comparative account of harm should also allow for cumulative bad states. What I mean by that is that the non-comparative account should be able to explain how someone can be harmed by being put into a bad state even if one is already in such a state. For example, suppose that a pedestrian has a toothache, harming him by putting him into a non-comparatively bad state (pain). And then suppose the person gets hit in the head by tree branch, leading to a headache that is worse than the toothache. Although the pedestrian was already in pain, the new pain should still count as a harm to him. Otherwise, inflicting non-comparatively bad states on people who already have a bit of the bad state would not be harmful. That would be an unacceptable conclusion.

Exactly what states are non-comparatively bad is hotly debated; some say ignorance, while others might suggest that pain is the only non-comparatively bad state to be in. I am more

sympathetic to the latter. I hold the Epicurean view that something cannot be bad for someone unless that person is capable of experiencing it. Fisher sums it up nicely in two principles:<sup>xviii</sup>

Experience Requirement I (ER I): An individual can be harmed by something only if he has an unpleasant experience as a result of it (either directly or indirectly).

Experience Requirement II (ER II): An individual can be harmed by something only if it is possible for him to have an unpleasant experience as a result of it (either directly or indirectly).

Unpleasant experiences require feelings, so this definition of harm requires feelings. Unlike Epicurus, I do not think that pain is the only intrinsic bad, but I think that the only intrinsic bads are negative feelings, which include pain. Using pain as the main example of an intrinsic bad will be sufficient for this paper though.

The non-comparative account of harm has a “problem” in that it cannot for the supposed harm of death. Death is not a non-comparatively bad state, so one can’t be harmed by being dead. This is certainly an issue for proponents of the non-comparative account who want to also keep death as a type of harm. Because of my controversial position, I need not worry about this particular problem since I do not think death is harmful, being that it cannot be experienced. If I can succeed in explaining that only things that can be experienced can be harmful, then the common intuition that death can be harmful need not carry much weight.

Both the comparative and non-comparative accounts of harm have some odd consequences. Both lead to cases where death is not harmful. This is truer for the non-comparative account though, since it states that death can never be harmful for the person who experiences it. The comparative account has larger overall flaws though, since it can’t account for overdetermined cases harm. It also operates on a plastic case by case basis, which initially seems like a good thing. However, it quickly runs into problems when one realizes that there is

no non-arbitrary way to determine whether someone is harmed in a particular case, since there are near-infinite possibilities in some cases. The non-comparative account can explain overdetermined cases of harm and does not need case-by-case comparisons, so although it might frustrate people who want to consider death a harm, it is altogether a better theory of harm.

Nagel argues strongly against the idea that harm has to be experienced to be bad. To support his argument, Nagel states that a betrayal that is unknown to the victim wouldn't be bad for the victim as long as he doesn't experience it. Nagel himself puts it much more eloquently: "It [the view that what you don't know can't hurt you] means that even if a man is betrayed by his friends, ridiculed behind his back, and despised by people who treat him politely to his face, none of it can be counted as a misfortune for him so long as he does not suffer as a result."<sup>xix</sup>

As Fisher and others have pointed out, this betrayal scenario provides a counter-example to ERI but not ERII. The betrayal can still be bad for the victim because it is possible for the victim to find out about it or at least experience the negative happenings that are caused by being betrayed. Thus Nagel has failed to prove ERII false. While Fisher acknowledges Nagel's failure, Fisher himself goes on to attempt a counterexample against ERII. Fisher uses a betrayal example very similar to Nagel's which differs only in that it has a fail-safe mechanism in the form of White. White is in place to prevent the supposed victim from ever learning about the betrayal if he ever gets close to discovering it. As it turns out though, White is not needed in the actual case because the "victim" never gets close to revealing the truth.

Fisher argues that Nagel's example and his own example are relevantly similar enough that it would be implausible to say that someone is harmed in the first case but not the second. However, I believe that Fisher has made a large, relevant difference between the two cases with his addition of White. In the original case, it is possible for one to find out about the betrayal and

be harmed by it. Actually one does not even need to find out about it, if there are negative experiential effects that affect the betrayal victim, then the betrayal victim has been harmed. Fisher's addition of White makes it impossible for the target of the betrayal to ever find out about it. Despite what Fisher may think, this actually supports ERII. It is only possible for someone to be harmed if he can have a negative experience of it, and White makes it impossible to have a negative experience. Fisher's attempt at a counterexample of ERII thus only provides an actual example of ERII, not an objection.

Nagel's argument for the harmfulness of death rests partially on him being a deprivation theorist. Nagel says that "if death is an evil at all, it cannot be because of its positive features, but only because of what it deprives us of."<sup>xx</sup> While Nagel is not a pure proponent of the comparative account of harm, since he thinks death is unconditionally bad, his views on deprivation are similar to those of more traditional comparativists. Death makes people worse off than they would have been since death "is an abrupt cancellation of indefinitely extensive possible goods."<sup>xxi</sup> And on the comparative account of harm, such harms are not dependent on feelings, because a lack of basic goods need not require any cognitive or affective experience of the lack. For comparativists, deprivation of goods can be intrinsically harmful.

I hold an asymmetrical view of goods and bads. While I believe that there are only psychological bads that are based on experiences and feelings, I think there are some non-psychological goods. What things are basic goods is quite contentious, and since this paper is about harm, and not the nature of basic goods, I am not going to argue about what things actually constitute basic goods. However a few concrete possibilities would help, so qualities such as knowledge, virtue, and autonomy might be basic intrinsic goods.



I think that while the deprivation of basic goods can be harmful, the deprivation is not always harmful. For example, not all deprivations are bad for people. If knowledge is an intrinsic good, it does not follow on the non-comparative account that I am harmed since I do not know all the possible facts about the universe. On the comparativist account, I also cannot be harmed by lacking that knowledge, assuming that it was never possible for me to ever know all the facts about the universe. This is true even in cases of where someone has some knowledge but loses it later. Suppose I had knowledge of a current tasteless pop song that played endlessly in my head, but then later forgot that song and never remembered it again. I do not think it follows, even on the comparative account, that I am harmed by losing knowledge about a terrible tune. If it is a deprivation, it is not a meaningful one. It might even be good for me overall. While one may object that only knowledge about certain things are valuable, this objection would lead to an even bigger problem. If only knowledge about certain things is good, then knowledge is not intrinsically good, since the goodness of knowledge would be dependent on the types of things that are known.

Even in cases where deprivation is harmful, the harmfulness is not intrinsic. Deprivation is only an intrinsic harm on the comparative account, and as I have argued earlier, the comparative account of harm is horribly flawed. On the non-comparative account, one can only be harmed if the deprivation puts one into an intrinsically bad state. Ignorance is not an intrinsically bad state, since one is not in a non-comparatively bad state in each scenario where one is ignorant. The “not having knowledge of the entire universe” is one example where ignorance is not non-comparatively bad. Ignorance tends to cause harm to people since it places them into intrinsically bad states such as pain, but this is an instrumental effect of being ignorant, not an intrinsic effect of ignorance itself.

One of the difficulties with the whole experiential version of harm is the role that intuition plays. For some people like me, it is intuitive that someone can only be harmed if they have the possibility of experiencing the harm. Most philosophers probably have the opposite intuition; there are things besides pain and other negative feelings that are intrinsically bad. A lot of the arguments against the experiential account rely almost solely on intuition.<sup>2</sup> Given this state of affairs, I doubt I can persuade everyone within the scope of this paper, but at the least I might be able to provide reasons for why someone might intuitively think that there are other intrinsic bad without there actually being such things.

In real life, negative events like betrayal or promise breaking are almost always experientially harmful to the victim who experiences them. It is also almost always possible to find out about such negative things; stipulations of so-and-so being impossible to discover are more the domains of thought experiments than real life. Since betrayals and promise breakings lead to negative effects all the time in real life, it is understandable why someone could come to believe that those things are intrinsically bad instead of instrumentally bad. After all, sometimes it just *seems* that betrayal is bad in itself. And intrinsic bads are fundamentally bad; they don't need an argument or explanation. Instead they are just stipulated as such. So it may be impossible to reconcile intuitions on what things are bad, since these may reflect unchangeable foundational philosophical differences.

### **Wrongness**

While harm is regarded as a bad thing, not all harms are wrong. Wrongness implies moral value (bad or negative moral value), but not all harms are morally relevant. This is true regardless of whether one takes a comparative or non-comparative account of harm. Take the

---

<sup>2</sup> For example, Nozick's Experience Machine argument relies on the premise that people would not want to plug into the machine. That is not wholly true, since it is quite possible some people could rationally will to plug in.

head injury example earlier. If the person is hit on the head by another person, then the victim is harmed. However the person would also be harmed, perhaps even to the same extent, if that person was hit on the head with a falling tree branch. And yet the general intuition is that the harm is only considered morally relevant in the first case, but not the second.

Someone may be in a bad state, such as pain, that is harmful to that person. How that person was harmed affects whether the harm is morally relevant or not. One can use the head injury example once again, where the pain is caused by a person in the first case and by a tree in the second case. In the first case the harm is morally wrong, since someone causes the pain. The second case has no morally relevant harm since there is no agent causing the harm.

As I am concerned with the relationship between wrongness and harm in this paper, not all harms will be relevant because not all harms have moral worth. Harm is only morally relevant when there is an agent causing the harm. While natural disasters such as earthquakes and tornadoes can be harmful, they can never be harmful in a morally relevant sense because they are not agents. For instance, if a tornado eradicates a town it would make sense to say that the tornado did harm to the people living in the town. It would be absurd, however, to suggest that the tornado is morally wrong or that it was wrong for the tornado to do that; a tornado is a force of nature, not an agent. Only people can be agents, so only people can cause morally relevant harm. It should be noted that by people I mean entities with the capacity for moral thought and action; this category includes but is not limited to human beings. Presumably any conscious being has the capacity for moral action.

Morality involves agents but not every harm involving agents necessitates a utilization of morality. Consider the case of Jones and Smith. Jones and Smith have both been waiting in line for flu shots, with Jones in front of Smith in virtue of arriving earlier. As it turns out, Jones

receives the last flu shot and Smith gets nothing. Lacking a flu shot, Smith gets sick. It would seem that according to the comparative account of harm that Smith is harmed by Jones being in line, since it prevents Smith from getting a flu shot and not getting sick. However, even though Jones purposely received the flu shot, it doesn't seem that Jones has harmed Smith in a morally relevant way, assuming that Jones did not get the shot in order to spite Smith. So even though Jones' action led to Smith being harmed, it would be absurd to suggest that Jones harmed Smith in a morally relevant sense.

For simplicity, I will call morally relevant harms 'moral harms'. This definition suggests that the harm caused is morally wrong, not that one's morals are being harmed (hence the use of quotations to differentiate it from this usage). The question of what things are morally harmful is debatable, so I will focus on relatively uncontroversial examples of 'moral harm'.

I will argue that all 'moral harm' is based on feelings, particularly pain. When I say pain I am referring to all categories of pain: physical pain, psychological pain, etc. Although not everything that is wrong and 'morally harmful' relies directly on feelings, everything in this category has to at least rely indirectly on feelings.

Even if death itself is not harmful, killing is still wrong. However, the common idea that killing is wrong because it is a deprivation of life is not an adequate explanation for the wrongness of killing. For example, suppose that an assassin kills two people, P1 and P2. P1 would have gone on to live a vastly pleasurable life, while P2 would have been completely miserable. (If one likes, one can replace the term "pleasurable" with whatever qualities the reader thinks are important to a good life. The specifics don't matter for the example). Intuition would dictate that the killing of P1 and P2 are about equally wrong, even though that P1 and P2 would go on to have vastly different life experiences. Even if the killing of P2 can be thought of as less

immoral than the killing of P1, the killing of P2 still seems wrong. Thus, the wrongness of the act killing can't be explained completely by the loss of the benefits of life.

Also, the wrongness of killing cannot be directly based on pain as a harm. Although most instances of killing do involve pain as a harm, it is easy to think of counterexamples where someone is killed rather painlessly, perhaps without even being aware of it. Dying peacefully in one's sleep would be an example of this. So the harm cannot be based directly on feelings. In order to establish a good relationship between the harm and the wrongness of killing, another foundation needs to be established.

Attempts to cause harm are important since the attempts are often morally wrong without actually being harmful. They are not harmful because the would-be victim is actually no worse off in either the comparative or state-based account of harm. In the comparative account, the person is not made worse off than he or she would have been otherwise since nothing negative happened. In the state-based account, that person is not put into a bad state since once again nothing happened. However, the attempt to harm is still wrong even if there is no actual harm being caused. The amount of wrongness of the attempt is dependent on the wrongness of the harmful act had the act been successful. More potentially harmful attempts (like attempting to torture someone) are seen as more wrong than less potentially harmful attempts (like attempting to steal \$5). So attempts to harm are counterfactually dependent on harm, even if by definition no harm actually occurs.

In conclusion, I have argued that feelings have an important relationship to moral wrongness because of the nature of harm. Harm is often explained by feelings, though it always has at least an indirect connection to them. Western morality tends to judge things to be morally wrong on the basis of harm, as opposed to authority, purity or other moral foundations. There are

differences between feelings and emotions in that emotions entail feelings but also have intentionality towards objects. Feelings are what give emotions their motivational weight. If I have correctly shown the relationship between feelings, harm, and wrongness, then feelings determine wrongness, at least in part. There are other factors, such as reason, at play in morality but my hypothesis is that feelings play an essential role in our understanding of morality.

---

<sup>i</sup> Solomon, Robert. (1976). The Passions: Emotions and the Meaning of Life. Hackett Publishing Company, Garden City NY, pg 97

<sup>ii</sup> MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social pain and physical pain. *Psychological Bulletin*, 131, 202–223.

<sup>iii</sup> Rose GA. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bull World Health Organ* 1962;27:645-58.

<sup>iv</sup> Hilgard, E.R., Morgan, A.H. & Macdonald, H. (1975). Pain and dissociation in the cold pressor test: a study of hypnotic analgesia with “hidden reports” through automatic key-pressing and automatic talking. *Journal of Abnormal Psychology*, 190, pp. 280–9.

<sup>v</sup> de Sousa, Ronald, "Emotion", *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2010/entries/emotion/>>.

<sup>vi</sup> James, William. (1884). What is an Emotion? *Mind*, 9: 188–205.

<sup>vii</sup> Schacter, Stanley, and Singer, Jerome. (1962). Cognitive, Social and Physiological Determinants of Emotional States,” *Psychological Review*, 69: 379–99.

<sup>viii</sup> Sinclair, R. C., Hoffman, C., Mark, M. M., Martin, L. L., & Pickering, T. L. (1994). Construct accessibility and the misattribution of arousal: Schachter and Singer revisited. *Psychological Science*, 5, 15-19.

<sup>ix</sup> Deigh, John. (1994). Cognitivism in the Theory of Emotions, *Ethics*, 104: 824–54.

<sup>x</sup> S.G. Anagnostaras, G.D. Gale and M.S. Fanselow. (2001). Hippocampus and contextual fear conditioning: recent controversies and advances. *Hippocampus* 11, 8–17.

<sup>xi</sup> Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.

<sup>xii</sup> Hanser, M. (2008). The Metaphysics of Harm. *Philosophy and Phenomenological Research* 77: 421-50.

<sup>xiii</sup> McMahan J. (2002). *The Ethics of Killing*. Problems at the Margins of Life. Oxford: Oxford University Press; Pg 115

<sup>xiv</sup> McMahan. *Ethics of Killing*. pg 111

<sup>xv</sup> Hanser *Metaphysics of Harm* pg 434

<sup>xvi</sup> Hanser, *Metaphysics of Harm* pg 435

<sup>xvii</sup> Hanser, *Metaphysics of Harm*, pg 426

<sup>xviii</sup> Fischer, J. M. (1997). Death, badness, and the impossibility of experience. *Journal of Ethics*, 1, 341–353.

<sup>xix</sup> Nagel, Thomas. (1970). Death. *Nous* 4: pg 76

<sup>xx</sup> Nagel, Death. Pg 74

<sup>xxi</sup> Nagel, Death. Pg 80

## References

S.G. Anagnostaras, G.D. Gale and M.S. Fanselow. (2001). Hippocampus and contextual fear conditioning: recent controversies and advances. *Hippocampus* 11, 8–17.

- 
- Cima M, Tonnaer F, Hauser M. (2010). Psychopaths know right from wrong but don't care. *Social Cognitive and Affective Neuroscience*; 5:59-67.
- Deigh, John. (1994). Cognitivism in the Theory of Emotions. *Ethics*, 104: 824–54.
- Fischer, J. M. (1997). Death, badness, and the impossibility of experience. *Journal of Ethics*, 1, 341–353.
- Gill, M. B. and Nichols, S. (2008). Sentimentalist pluralism: Moral psychology and philosophical ethics. *Philosophical Issues*, 18: 143–163.
- Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong, (Ed.), *Moral psychology* (Vol. 3). The neuroscience of morality: emotion, disease, and development. Cambridge, MA: MIT Press.
- Greene, J. (2003). From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology? *Nat Rev Neurosci* 4, 846-849.
- Greene, J., and Haidt, J. (2002). How (and where) does moral judgment work? *Trends Cogn Sci* 6, 517-523.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., and Cohen, J. D. (2001). An fmri investigation of emotional engagement in moral judgment. *Science* 293, 2105-2108.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108, 814-834.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998–1002.
- Haidt, J., Bjorklund, F., and Murphy, S. (2000). Moral dumbfounding: When intuition finds no reason. (unpublished manuscript, University of Virginia).
- Haidt, J., Koller, S. H., and Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *J Pers Soc Psychol* 65, 613-628.

---

Hanser, M. (2008). "The Metaphysics of Harm." *Philosophy and Phenomenological Research* 77: 421-50.

Hilgard, E.R., Morgan, A.H. & Macdonald, H. (1975). Pain and dissociation in the cold pressor test: a study of hypnotic analgesia with "hidden reports" through automatic key-pressing and automatic talking. *Journal of Abnormal Psychology*, 190, 280–9.

James, William. (1884). What is an Emotion? *Mind*, 9: 188–205.

MacDonald, G., & Leary, M. R. (2005). Why does social exclusion hurt? The relationship between social pain and physical pain. *Psychological Bulletin*, 131, 202–223.

McMahan, Jeff. (1988). "Death and the Value of Life." *Ethics* 99: 32–61.

McMahan J. (2002). *The Ethics of Killing. Problems at the Margins of Life*. Oxford: Oxford University Press;

Mendez, M., Anderson, E., Shapira, J.S. (2005). An investigation of moral judgment in frontotemporal dementia. *Cognitive and Behavioral Neurology*, 18, 193–197.

Nagel, Thomas. (1970). Death. *Nous* 4: 73–80

Nozick R. (1971). *The Experience Machine*. *Moral Philosophy: A Reader*, ed. Louis Pojman, pp. 124-191. Hackett Publishing Company, Inc., Indianapolis.

Rose, GA. (1962). The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bull World Health Organ*; 27:645-58.

Schacter, Stanley, and Singer, Jerome. (1962). Cognitive, Social and Physiological Determinants of Emotional States. *Psychological Review*, 69: 379–99.

Schnall, S., Haidt, J., and Clore, G. (2004). Irrelevant disgust makes moral judgment more severe, for those who listen to their bodies.



---

Sinclair, R. C., Hoffman, C., Mark, M. M., Martin, L. L., & Pickering, T. L. (1994). Construct accessibility and the misattribution of arousal: Schachter and Singer revisited. *Psychological Science*, 5, 15-19.

Solomon, Robert. (1976). The Passions: Emotions and the Meaning of Life. Hackett Publishing Company, Garden City NY.

de Sousa, Ronald, "Emotion", *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), <http://plato.stanford.edu/archives/spr2010/entries/emotion/>

Thomson, Judith Jarvis. (2011). More on the Metaphysics of Harm. *Philosophy and Phenomenological Research* 82: 436–458.